Manifold Sampling for Nonconvex Piecewise Continuously Differentiable Functions

Jeffrey Larson

Stefan Wild, Kamil Khan, Matt Menickelly

Argonne National Laboratory

July 12, 2016

Problem Statement

We are interested in solving the problem:

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) \equiv h(F(x))$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$, $h : \mathbb{R}^m \to \mathbb{R}$,

Problem Statement

We are interested in solving the problem:

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) \equiv h(F(x))$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$, $h : \mathbb{R}^m \to \mathbb{R}$, and

- h is nonsmooth, piecewise linear, but has a known structure (cheap to evaluate)

Problem Statement

We are interested in solving the problem:

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) \equiv h(F(x))$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$, $h : \mathbb{R}^m \to \mathbb{R}$, and

- h is nonsmooth, piecewise linear, but has a known structure (cheap to evaluate)
- F is smooth, nonlinear, but has a relatively unknown structure (expensive to evaluate)

Problem Statement

We are interested in solving the problem:

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) \equiv h(F(x))$$

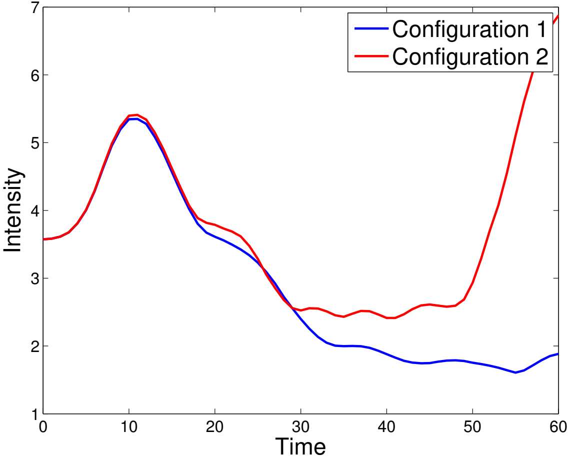
where $F : \mathbb{R}^n \to \mathbb{R}^m$, $h : \mathbb{R}^m \to \mathbb{R}$, and

- h is nonsmooth, piecewise linear, but has a known structure (cheap to evaluate)
- F is smooth, nonlinear, but has a relatively unknown structure (expensive to evaluate)

Though h is piecewise linear, F being nonlinear implies f can be nonlinear

Laser pulse propagating in a plasma channel

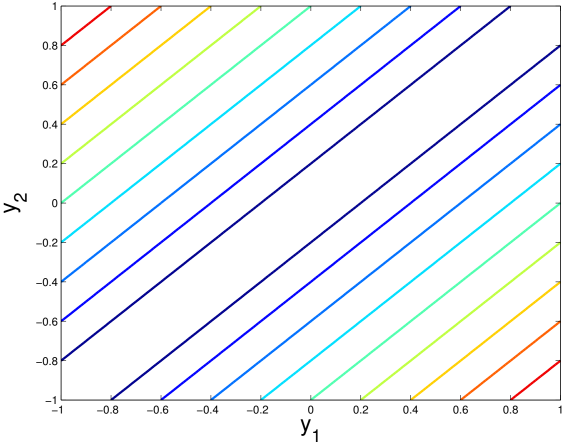
Want to determine the plasma channel properties so the maximum difference in the laser intensity during propagation is minimized.



$$f(x) = \max\{F_i(x)\} - \min\{F_i(x)\}$$

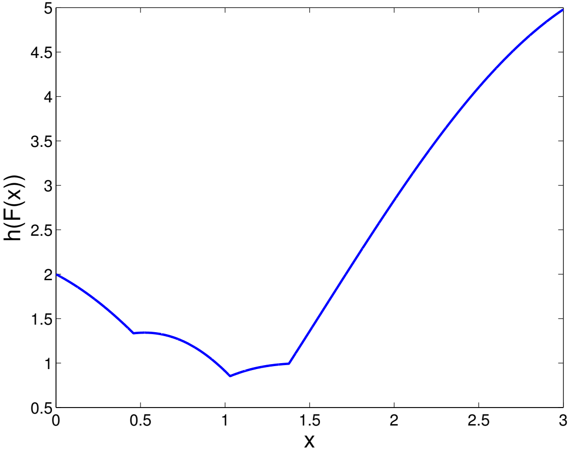
Formulation

$$h(y) = \max\{y\} - \min\{y\}$$



Formulation

$$h(F(x)) = \max\{\sin(2x)+1, \cos(2x), x\} - \min\{\sin(2x)+1, \cos(2x), x\}$$



Nonsmooth Optimization

Smooth optimization

Nonsmooth optimization

Nonsmooth Optimization

Smooth optimization

- A descent direction, $-\nabla f(x)$, exists everywhere

Nonsmooth optimization

Nonsmooth Optimization

Smooth optimization

- A descent direction, $-\nabla f(x)$, exists everywhere
- $\nabla f(x) = 0$ is a necessary optimality condition

Nonsmooth optimization

Nonsmooth Optimization

Smooth optimization

- A descent direction, $-\nabla f(x)$, exists everywhere
- $\nabla f(x) = 0$ is a necessary optimality condition
- Difference approximation can approximate $\nabla f(x)$

Nonsmooth optimization

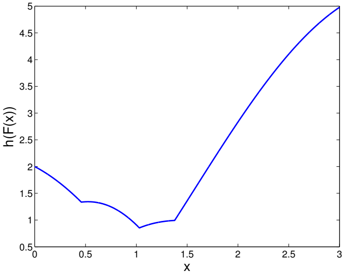
Nonsmooth Optimization

Smooth optimization

- A descent direction, $-\nabla f(x)$, exists everywhere
- $\nabla f(x) = 0$ is a necessary optimality condition
- Difference approximation can approximate $\nabla f(x)$

Nonsmooth optimization

- Gradient may not exist at every point



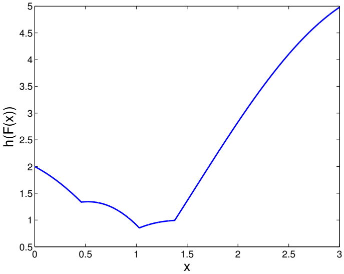
Nonsmooth Optimization

Smooth optimization

- A descent direction, $-\nabla f(x)$, exists everywhere
- $\nabla f(x) = 0$ is a necessary optimality condition
- Difference approximation can approximate $\nabla f(x)$

Nonsmooth optimization

- Gradient may not exist at every point
- Gradient often does not exist at the optimal point



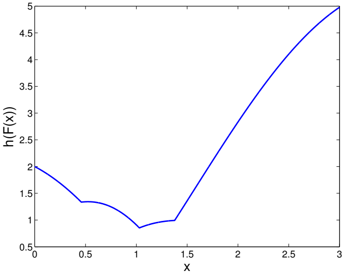
Nonsmooth Optimization

Smooth optimization

- A descent direction, $-\nabla f(x)$, exists everywhere
- $\nabla f(x) = 0$ is a necessary optimality condition
- Difference approximation can approximate $\nabla f(x)$

Nonsmooth optimization

- Gradient may not exist at every point
- Gradient often does not exist at the optimal point
- Difference approximation of $\nabla f(x)$ can be very bad



Nonsmooth Optimization

Smooth optimization

- A descent direction, $-\nabla f(x)$, exists everywhere
- $\nabla f(x) = 0$ is a necessary optimality condition
- Difference approximation can approximate $\nabla f(x)$

Nonsmooth optimization

- Gradient may not exist at every point
- Gradient often does not exist at the optimal point
- Difference approximation of $\nabla f(x)$ can be very bad

Just running a smooth algorithm on a nonsmooth problem may not converge, or may converge to a nonstationary point.

A generalized derivative

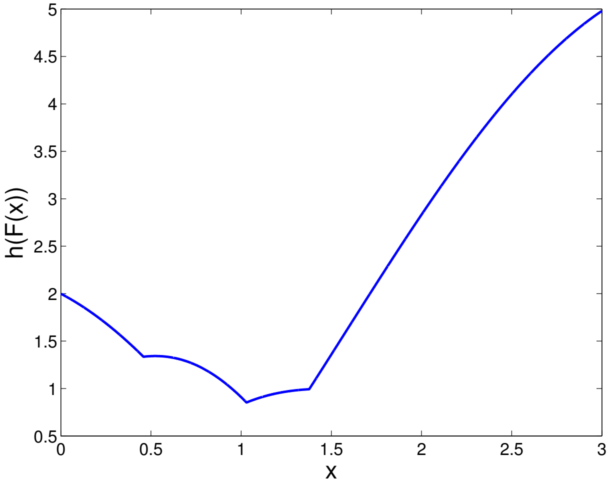
Definition

For locally Lipschitz continuous functions $f : \mathbb{R}^n \to \mathbb{R}$, the *generalized Clarke subgradient* of f at a point $x \in \mathbb{R}^n$ is:

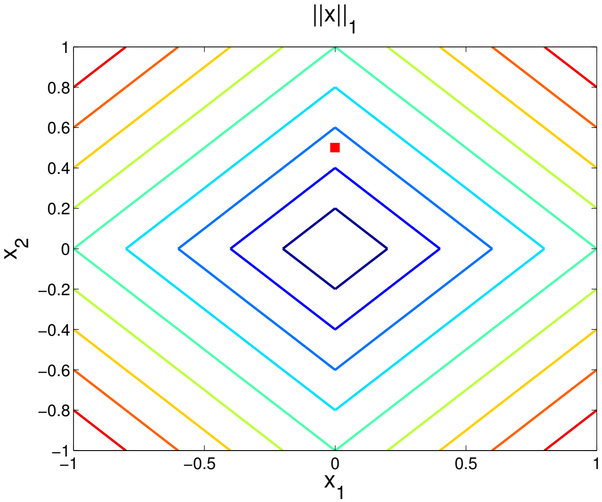
$$\partial f(x) = \operatorname{conv}\left\{\xi \in \mathbb{R}^n : \xi = \lim_{x^i \to x} \nabla f(x^i) \text{ and } \nabla f(x^i) \text{ exists at all } x^i\right\}$$

where $\operatorname{conv}(\cdot)$ denotes the convex hull of a set.

A generalized derivative



A generalized derivative



$$\partial \left\| [0, 0.5] \right\|_1 = \operatorname{conv} \{ [1, 1], [-1, 1] \}$$

A generalized derivative

- If f is locally Lipschitz continuous and differentiable at x, $\nabla f(x) \in \partial f(x)$.

A generalized derivative

- If f is locally Lipschitz continuous and differentiable at x, $\nabla f(x) \in \partial f(x)$.

- If f is locally Lipschitz continuous and continuously differentiable at x, $\nabla f(x) = \partial f(x)$.

A generalized derivative

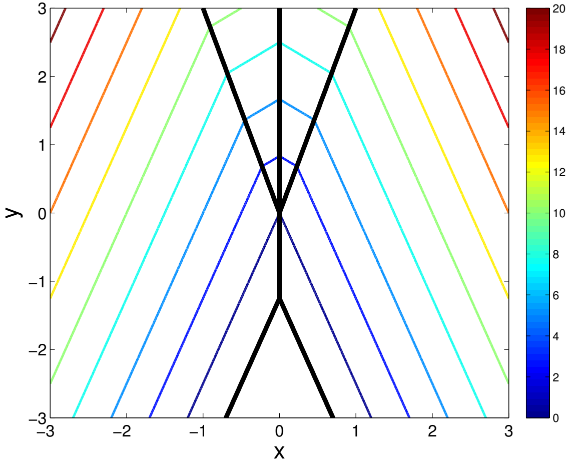
- If f is locally Lipschitz continuous and differentiable at x, $\nabla f(x) \in \partial f(x)$.
- If f is locally Lipschitz continuous and continuously differentiable at x, $\nabla f(x) = \partial f(x)$.
- $0 \in \partial f(x)$ is a necessary optimality condition for locally Lipschitz continuous f.

Steepest descent can have trouble. . .

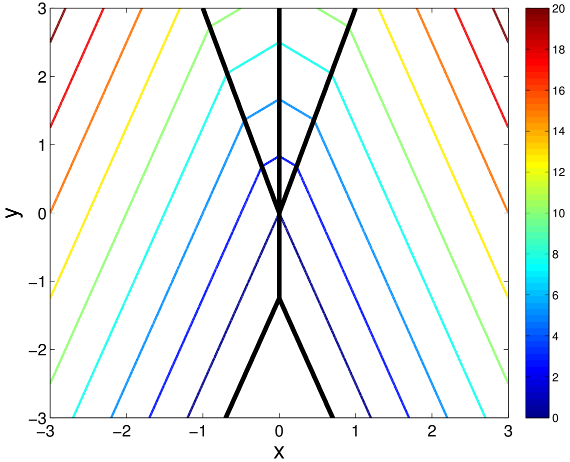
$$f(x,y) = \max\left\{-\frac{5}{2}, \pm 2x + 3y, \pm 5x + 2y\right\}$$

Piecewise affine, convex function.

Steepest descent can have trouble...

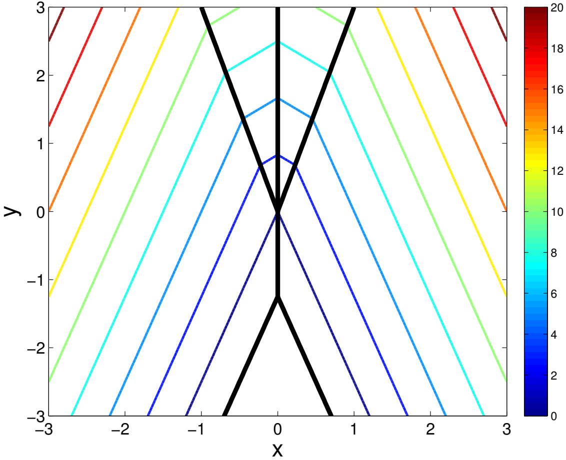


Steepest descent can have trouble...



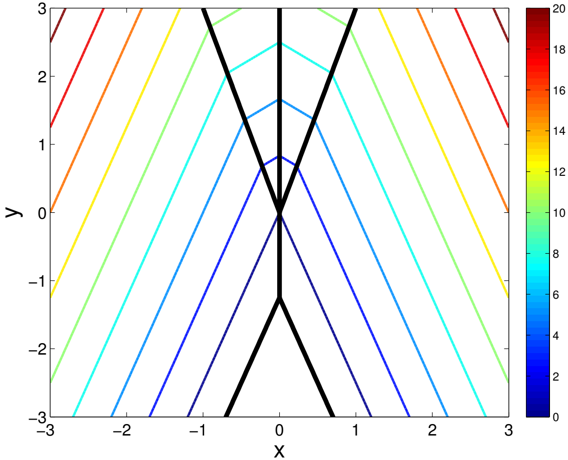
Steepest descent with exact line search started from any point on the diagonal lines will converge to $(0, 0)$.

Steepest descent can have trouble...



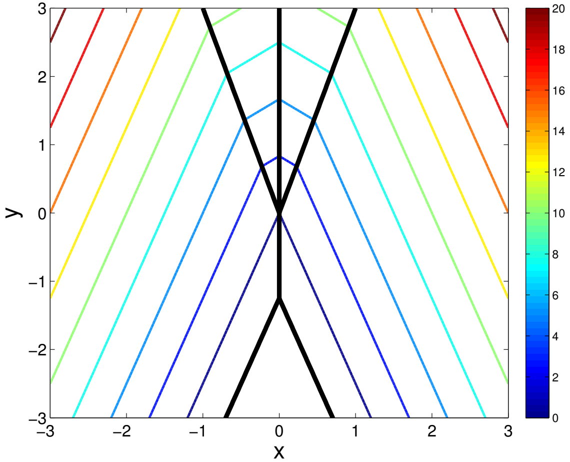
Steepest descent with exact line search started from any point above the diagonal lines will converge to $(0, 0)$.

Steepest descent can have trouble...



Steepest descent with inexact line search (producing points close to the diagonal) started from any point above the diagonal lines will converge to $(0, 0)$.

Steepest descent can have trouble...



Descent (with directions sufficiently close to steepest) with inexact line search (producing points close to the diagonal) started from any point above the diagonal lines will converge to $(0, 0)$.

Subgradient methods

Approaches for nonsmooth optimization

$$x^{k+1} = x^k + \alpha_k \xi^k$$

where ξ^k is some element in $\partial f(x^k)$,

Subgradient methods

Approaches for nonsmooth optimization

$$x^{k+1} = x^k + \alpha_k \xi^k$$

where ξ^k is some element in $\partial f(x^k)$, $\displaystyle\sum_{k=0}^{\infty} \alpha_k = \infty$,

Subgradient methods

Approaches for nonsmooth optimization

$$x^{k+1} = x^k + \alpha_k \xi^k$$

where ξ^k is some element in $\partial f(x^k)$, $\displaystyle\sum_{k=0}^{\infty} \alpha_k = \infty$, and $\displaystyle\sum_{k=0}^{\infty} \alpha_k^2 = \infty$.

Subgradient methods

Approaches for nonsmooth optimization

$$x^{k+1} = x^k + \alpha_k \xi^k$$

where ξ^k is some element in $\partial f(x^k)$, $\sum_{k=0}^{\infty} \alpha_k = \infty$, and $\sum_{k=0}^{\infty} \alpha_k^2 = \infty$.

- Not a descent method because the step sizes are fixed.

Subgradient methods

Approaches for nonsmooth optimization

$$x^{k+1} = x^k + \alpha_k \xi^k$$

where ξ^k is some element in $\partial f(x^k)$, $\displaystyle\sum_{k=0}^{\infty} \alpha_k = \infty$, and $\displaystyle\sum_{k=0}^{\infty} \alpha_k^2 = \infty$.

- Not a descent method because the step sizes are fixed.
- Not a descent method because ξ^k may not be a descent direction.

Gradient Sampling

Approaches for nonsmooth optimization

Theorem (Rademacher)

If $S \subset \mathbb{R}^n$ is open and $f : S \to \mathbb{R}$ is locally Lipschitz on S, then f is differentiable almost everywhere on S.

Gradient Sampling

1. Approximate $\partial f(x^k)$ by sampling $m \geq n+1$ points $x^{k,j}$ in $\mathcal{B}(x^k, \epsilon_k)$. Set

$$G^k = \text{conv}\left\{\nabla f(x^{k,1}), \dots, \nabla f(x^{k,m})\right\}$$

Gradient Sampling

1. Approximate $\partial f(x^k)$ by sampling $m \geq n+1$ points $x^{k,j}$ in $\mathcal{B}(x^k, \epsilon_k)$. Set

$$G^k = \operatorname{conv}\left\{\nabla f(x^{k,1}), \ldots, \nabla f(x^{k,m})\right\}$$

2. Set ξ^k to be the minimum norm element in G^k.

Gradient Sampling

Approaches for nonsmooth optimization

1. Approximate $\partial f(x^k)$ by sampling $m \geq n+1$ points $x^{k,j}$ in $\mathcal{B}(x^k, \epsilon_k)$. Set

$$G^k = \operatorname{conv}\left\{\nabla f(x^{k,1}), \ldots, \nabla f(x^{k,m})\right\}$$

2. Set ξ^k to be the minimum norm element in G^k.
3. Set α_k to be the smallest power s of $\gamma \in (0,1)$ satisfying

$$f(x^k + \gamma^s \xi^k) < f(x^k) - \beta\gamma^s \left\|\xi^k\right\|$$

Gradient Sampling

Approaches for nonsmooth optimization

1. Approximate $\partial f(x^k)$ by sampling $m \geq n+1$ points $x^{k,j}$ in $\mathcal{B}(x^k, \epsilon_k)$. Set
$$G^k = \operatorname{conv}\left\{\nabla f(x^{k,1}), \ldots, \nabla f(x^{k,m})\right\}$$
2. Set ξ^k to be the minimum norm element in G^k.
3. Set α_k to be the smallest power s of $\gamma \in (0,1)$ satisfying
$$f(x^k + \gamma^s \xi^k) < f(x^k) - \beta\gamma^s \left\|\xi^k\right\|$$
4. If $\nabla f(x^k + \alpha_k \xi^k)$ exists, $x^{k+1} = x^k + \alpha_k \xi^k$.
   Else, find a point in $\hat{x} \in \mathcal{B}(x^k, \epsilon_k)$ satisfying
$$f(\hat{x}^k + \gamma^s \xi^k) < f(x^k) - \beta\alpha_k \left\|\xi^k\right\|$$
   and set $x^{k+1} = \hat{x}^k + \alpha_k \xi^k$.

Gradient Sampling

Approaches for nonsmooth optimization

1. Approximate $\partial f(x^k)$ by sampling $m \geq n+1$ points $x^{k,j}$ in $\mathcal{B}(x^k, \epsilon_k)$. Set
$$G^k = \operatorname{conv}\left\{\nabla f(x^{k,1}), \ldots, \nabla f(x^{k,m})\right\}$$
2. Set ξ^k to be the minimum norm element in G^k.
3. Set α_k to be the smallest power s of $\gamma \in (0,1)$ satisfying
$$f(x^k + \gamma^s \xi^k) < f(x^k) - \beta\gamma^s \left\|\xi^k\right\|$$
4. If $\nabla f(x^k + \alpha_k \xi^k)$ exists, $x^{k+1} = x^k + \alpha_k \xi^k$.
Else, find a point in $\hat{x} \in \mathcal{B}(x^k, \epsilon_k)$ satisfying
$$f(\hat{x}^k + \gamma^s \xi^k) < f(x^k) - \beta\alpha_k \left\|\xi^k\right\|$$
and set $x^{k+1} = \hat{x}^k + \alpha_k \xi^k$.

- Iterates must not be at points of nondifferentiability

Gradient Sampling

Approaches for nonsmooth optimization

1. Approximate $\partial f(x^k)$ by sampling $m \geq n+1$ points $x^{k,j}$ in $\mathcal{B}(x^k, \epsilon_k)$. Set
$$G^k = \text{conv}\left\{\nabla f(x^{k,1}), \ldots, \nabla f(x^{k,m})\right\}$$
2. Set ξ^k to be the minimum norm element in G^k.
3. Set α_k to be the smallest power s of $\gamma \in (0,1)$ satisfying
$$f(x^k + \gamma^s \xi^k) < f(x^k) - \beta\gamma^s \left\|\xi^k\right\|$$
4. If $\nabla f(x^k + \alpha_k \xi^k)$ exists, $x^{k+1} = x^k + \alpha_k \xi^k$.
Else, find a point in $\hat{x} \in \mathcal{B}(x^k, \epsilon_k)$ satisfying
$$f(\hat{x}^k + \gamma^s \xi^k) < f(x^k) - \beta\alpha_k \left\|\xi^k\right\|$$
and set $x^{k+1} = \hat{x}^k + \alpha_k \xi^k$.

- Iterates must not be at points of nondifferentiability
- A lot of sampling may be required

Trust region methods

Smooth case

1. Build a model m_k of f at x^k, for example

$$m_k(p) = f(x^k) + \nabla f(x^k)^T p + \frac{1}{2} p^T \nabla^2 f(x^k) p$$

Trust region methods

Smooth case

1. Build a model m_k of f at x^k, for example

$$m_k(p) = f(x^k) + \nabla f(x^k)^T p + \frac{1}{2} p^T \nabla^2 f(x^k) p$$

2. Find s^k that minimizes m_k subject to $\left\| s^k \right\| \leq \Delta_k$.

Trust region methods

Smooth case

1. Build a model m_k of f at x^k, for example

$$m_k(p) = f(x^k) + \nabla f(x^k)^T p + \frac{1}{2} p^T \nabla^2 f(x^k) p$$

2. Find s^k that minimizes m_k subject to $\|s^k\| \leq \Delta_k$.
3. Evaluate

$$\rho_k = \frac{f(x^k) - f(x^k + s^k)}{m(x^k) - m(x^k + s^k)}$$

Trust region methods

Smooth case

1. Build a model m_k of f at x^k, for example

$$m_k(p) = f(x^k) + \nabla f(x^k)^T p + \frac{1}{2} p^T \nabla^2 f(x^k) p$$

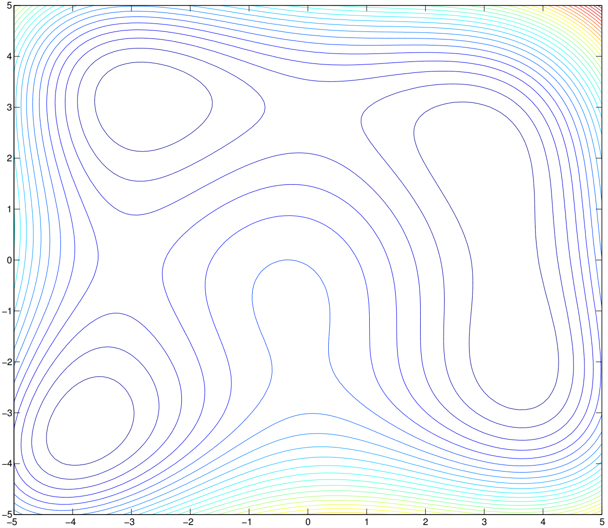
2. Find s^k that minimizes m_k subject to $\left\| s^k \right\| \leq \Delta_k$.
3. Evaluate

$$\rho_k = \frac{f(x^k) - f(x^k + s^k)}{m(x^k) - m(x^k + s^k)}$$

4. If $\rho_k > \eta > 0$, $x^{k+1} = x^k + s^k$, $\Delta_{k+1} = \gamma_{\mathrm{inc}} \Delta_k$.
   Else $x^{k+1} = x^k$, $\Delta_{k+1} = \gamma_{\mathrm{dec}} \Delta_k$.

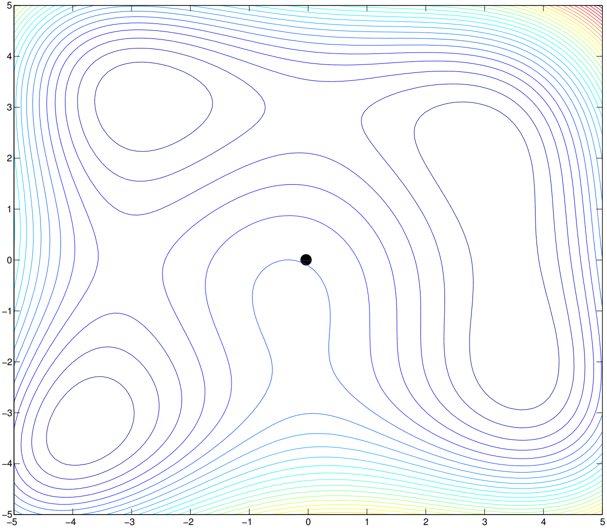
Trust region methods

Smooth case



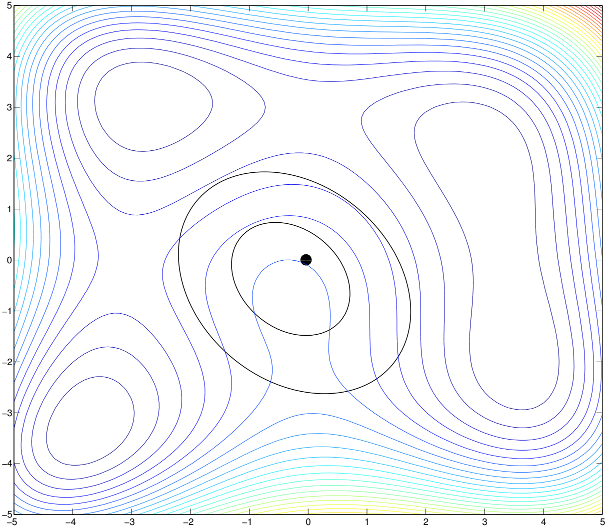
Trust region methods

Smooth case



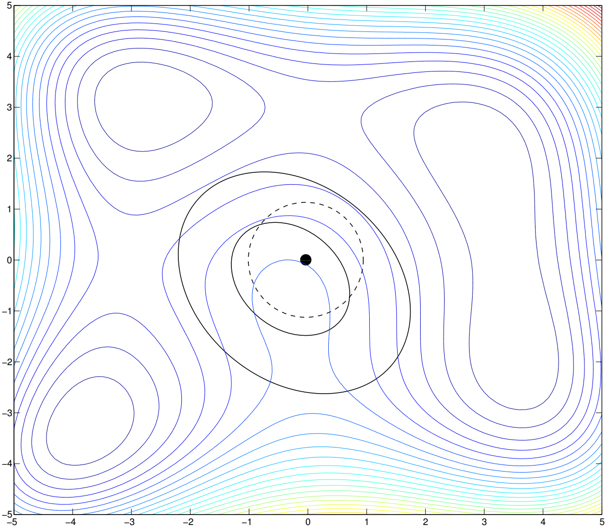
Trust region methods

Smooth case



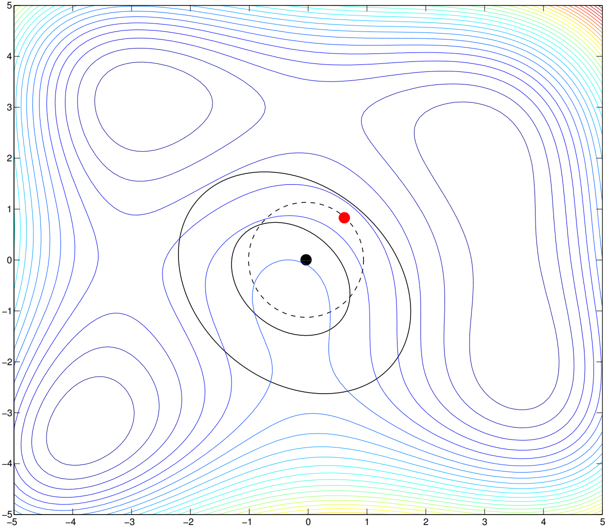
Trust region methods

Smooth case



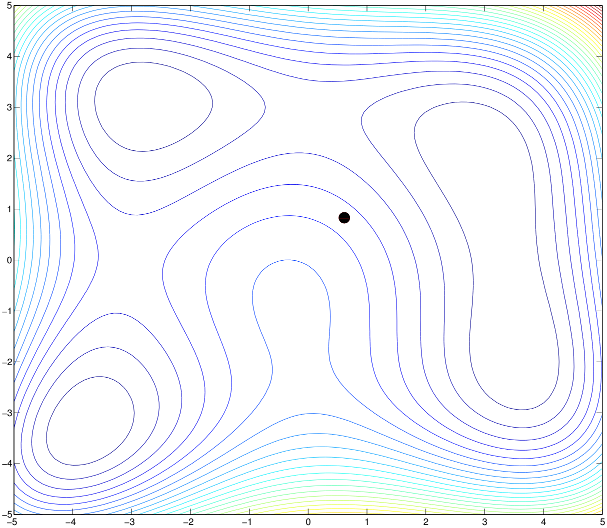
Trust region methods

Smooth case



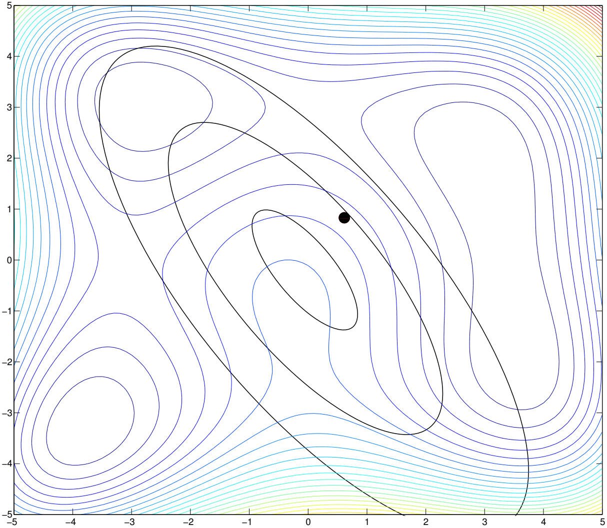
Trust region methods

Smooth case



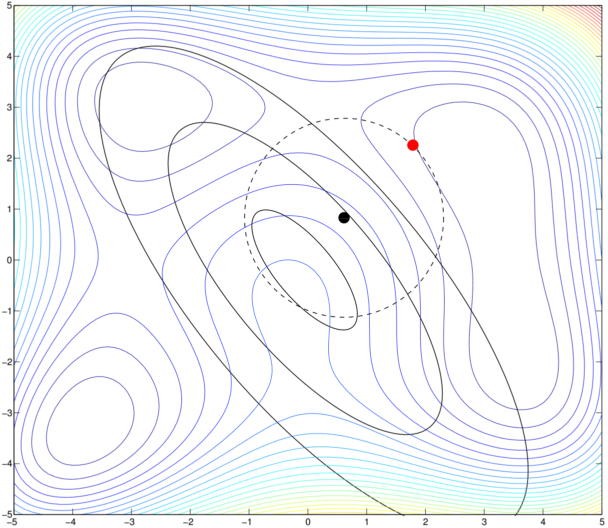
Trust region methods

Smooth case



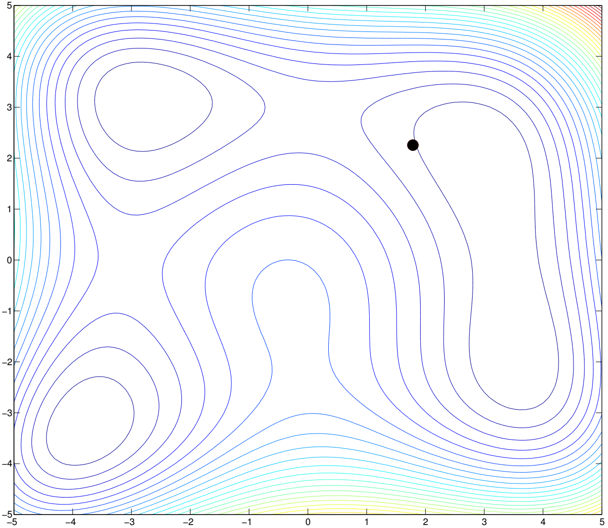
Trust region methods

Smooth case



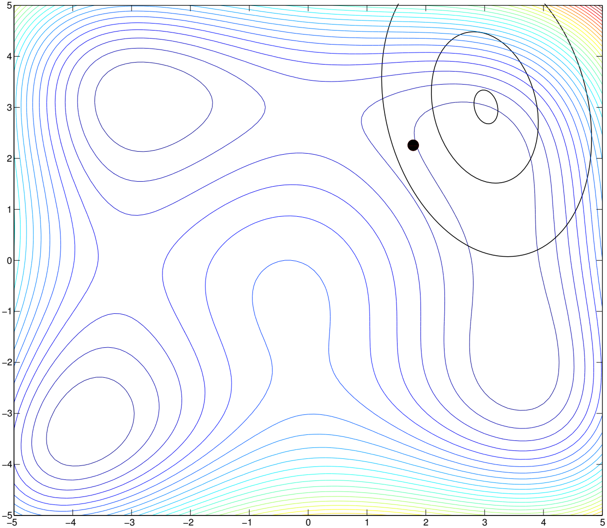
Trust region methods

Smooth case



Trust region methods

Smooth case



Traditional ρ test

Smooth Case

A model m sufficiently approximates f near x if

$$\begin{aligned} |f(x+s) - m(x+s)| &\le c_1 \Delta^2 \ \forall s \in \mathcal{B}(0,\Delta) \\ \|\nabla f(x+s) - \nabla m(x+s)\| &\le c_2 \Delta \ \forall s \in \mathcal{B}(0,\Delta), \end{aligned}$$

with c_1 and c_2 independent of Δ and x.

Traditional ρ test

Smooth Case

A model m sufficiently approximates f near x if

$$|f(x+s) - m(x+s)| \leq c_1 \Delta^2 \; \forall s \in \mathcal{B}(0, \Delta)$$
$$\|\nabla f(x+s) - \nabla m(x+s)\| \leq c_2 \Delta \; \forall s \in \mathcal{B}(0, \Delta),$$

with c_1 and c_2 independent of Δ and x.

$$\rho_k = \frac{f(x^k) - f(x^k + s^k)}{m(x^k) - m(x^k + s^k)}$$

Traditional ρ test

Smooth Case

A model m sufficiently approximates f near x if

$$|f(x+s) - m(x+s)| \leq c_1\Delta^2 \ \forall s \in \mathcal{B}(0,\Delta)$$
$$\|\nabla f(x+s) - \nabla m(x+s)\| \leq c_2\Delta \ \forall s \in \mathcal{B}(0,\Delta),$$

with c_1 and c_2 independent of Δ and x.

We want

$$\left|\frac{f(x^k) - f(x^k+s^k)}{m(x^k) - m(x^k+s^k)} - 1\right| \leq c\Delta_k$$

Traditional ρ test

Smooth Case

A model m sufficiently approximates f near x if

$$\begin{aligned} |f(x+s) - m(x+s)| &\leq c_1\Delta^2 \;\forall s \in \mathcal{B}(0,\Delta) \\ \|\nabla f(x+s) - \nabla m(x+s)\| &\leq c_2\Delta \;\forall s \in \mathcal{B}(0,\Delta), \end{aligned}$$

with c_1 and c_2 independent of Δ and x.

We want

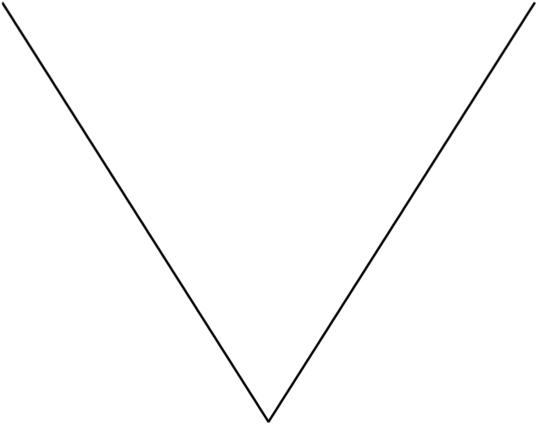
$$\left|\frac{f(x^k) - m(x^k) + m(x^k + s^k) - f(x^k + s^k)}{m(x^k) - m(x^k + s^k)}\right| \approx \frac{c_1\Delta_k^2 + c_1\Delta_k^2}{\|\nabla m(x)\|\,\Delta_k} \approx c_3\Delta_k$$

A new ρ test

Composite nonsmooth case

For nonsmooth functions, we do not get this.
If x^k and $x^k + s^k$ are on different sides of the absolute value kink,

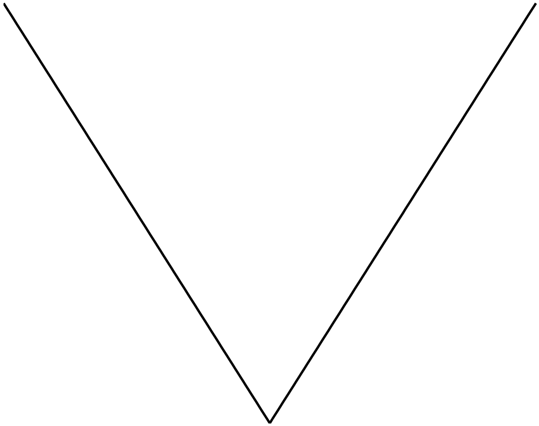
$$\|\nabla f(x+s) - \nabla m(x+s)\| = 2$$



A new ρ test

Composite nonsmooth case

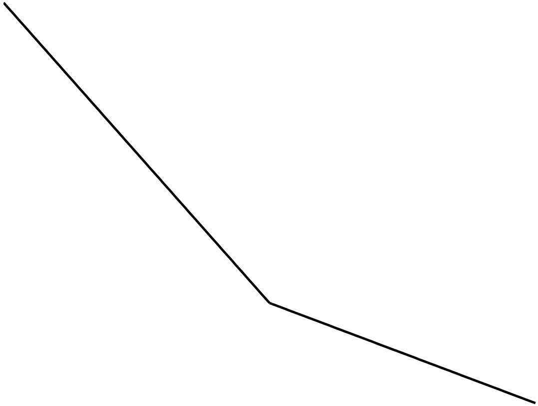
But if we include information at both x^k and $x^k + s^k$ when deciding on s^k, then we can ensure our model accurately approximates f on $\operatorname{conv}\left\{x^k, x^k + s^k\right\}$.



A new ρ test

Composite nonsmooth case

But if we include information at both x^k and $x^k + s^k$ when deciding on s^k, then we can ensure our model accurately approximates f on $\operatorname{conv}\left\{x^k, x^k + s^k\right\}$.



A new ρ test

Composite nonsmooth case

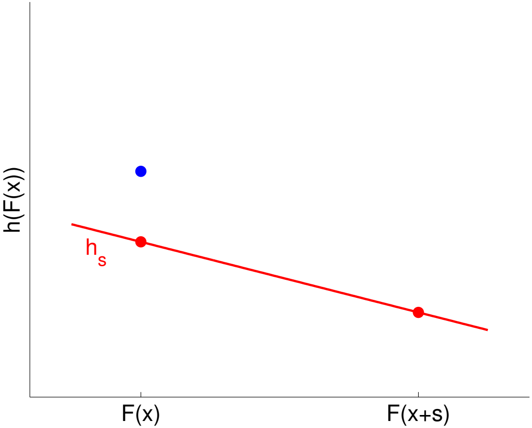
Let h_x and h_s be the affine functions active at $F(x^k)$ and $F(x^k + s^k)$, respectively. Then there are three cases.

A new ρ test

Composite nonsmooth case

Let h_x and h_s be the affine functions active at $F(x^k)$ and $F(x^k + s^k)$, respectively. Then there are three cases.

$$h_x(F(x)) \geq h_s(F(x))$$

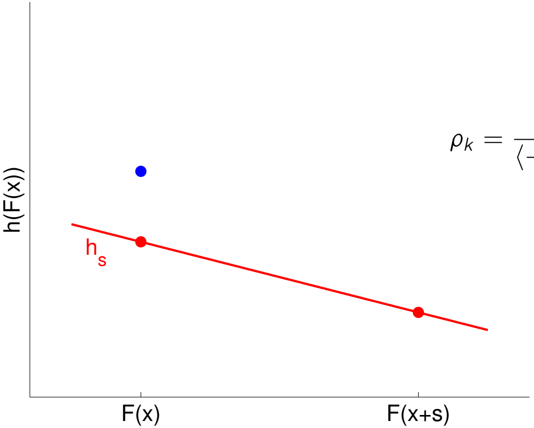


A new ρ test

Composite nonsmooth case

Let h_x and h_s be the affine functions active at $F(x^k)$ and $F(x^k + s^k)$, respectively. Then there are three cases.

$$h_x(F(x)) \geq h_s(F(x))$$



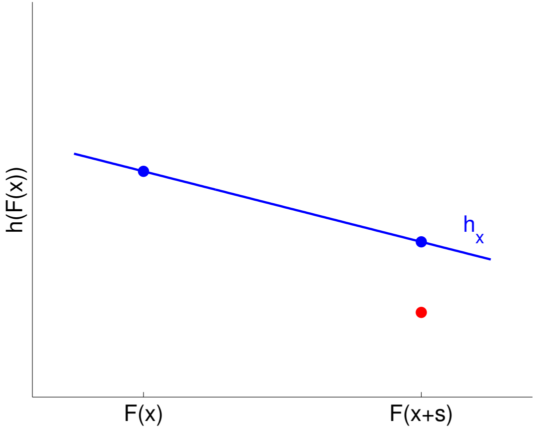
$$\rho_k = \frac{h_s(F(x^k)) - f(x^k + s^k)}{\langle -s^k, \nabla M(x^k) \nabla h_s(F(x^k)) \rangle}.$$

A new ρ test

Composite nonsmooth case

Let h_x and h_s be the affine functions active at $F(x^k)$ and $F(x^k + s^k)$, respectively. Then there are three cases.

$$h_x(F(x+s)) \geq h_s(F(x+s))$$

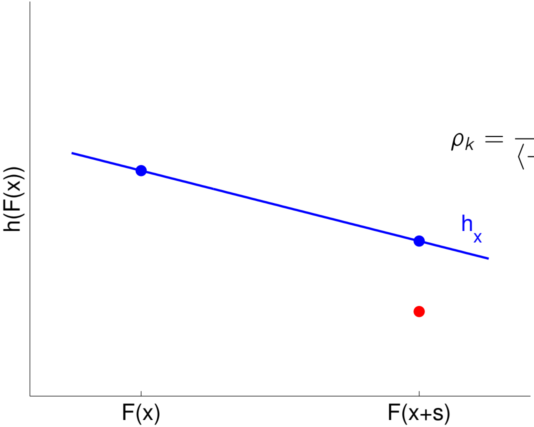


A new ρ test

Composite nonsmooth case

Let h_x and h_s be the affine functions active at $F(x^k)$ and $F(x^k + s^k)$, respectively. Then there are three cases.

$$h_x(F(x+s)) \geq h_s(F(x+s))$$



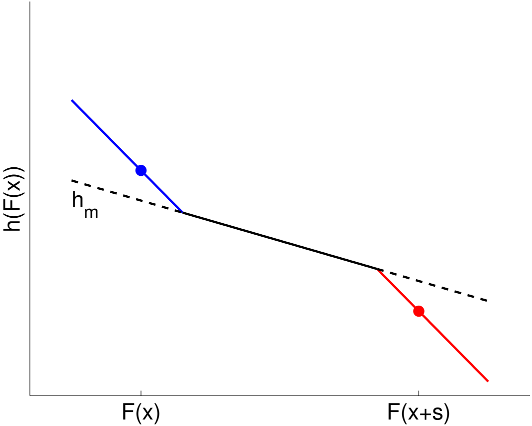
$$\rho_k = \frac{f(x^k) - h_x(F(x^k + s^k))}{\langle -s^k, \nabla M(x^k) \nabla h_x(F(x^k)) \rangle}.$$

A new ρ test

Composite nonsmooth case

Let h_x and h_s be the affine functions active at $F(x^k)$ and $F(x^k + s^k)$, respectively. Then there are three cases.

$$h_x(F(x)) \leq h_s(F(x)) \text{ and } h_x(F(x)) \leq h_s(F(x))$$

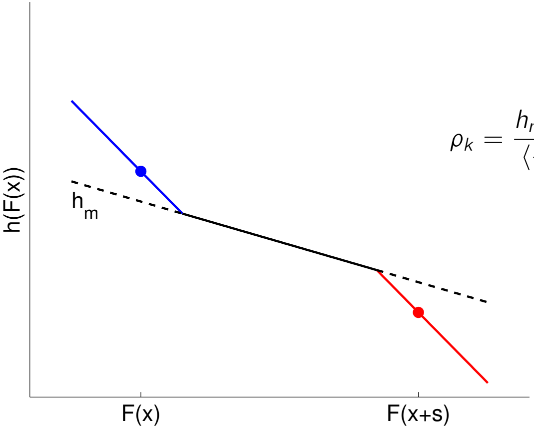


A new ρ test

Composite nonsmooth case

Let h_x and h_s be the affine functions active at $F(x^k)$ and $F(x^k + s^k)$, respectively. Then there are three cases.

$h_x(F(x)) \le h_s(F(x))$ and $h_x(F(x)) \le h_s(F(x))$



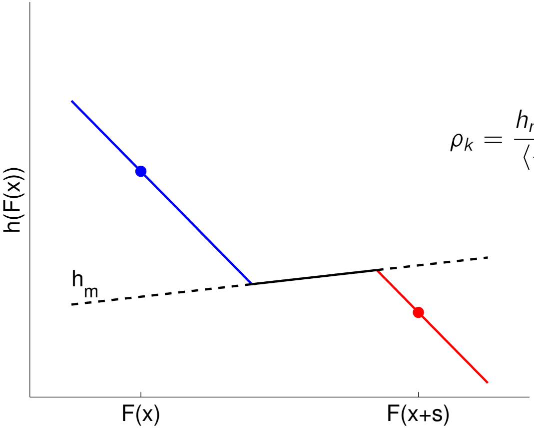
$$\rho_k = \frac{h_m(F(x^k)) - h_m(F(x^k + s^k))}{\langle -s^k, \nabla M(x^k) \nabla h_m(F(x^k)) \rangle}.$$

A new ρ test

Composite nonsmooth case

Let h_x and h_s be the affine functions active at $F(x^k)$ and $F(x^k + s^k)$, respectively. Then there are three cases.

$$h_x(F(x)) \leq h_s(F(x)) \text{ and } h_x(F(x)) \leq h_s(F(x))$$



$$\rho_k = \frac{h_m(F(x^k)) - h_m(F(x^k + s^k))}{\langle -s^k, \nabla M(x^k) \nabla h_m(F(x^k)) \rangle}.$$

Trust region methods

Composite nonsmooth case

1. Build a model $m_k^{F_i}$ of F_i at x^k.

Trust region methods

Composite nonsmooth case

1. Build a model $m_k^{F_i}$ of F_i at x^k.

2. Use ∇m^{F_i} to form $\nabla M(x)$ and build a set of generators G^k. Set ξ^k to be the minimum norm element in G^k.

Generator set

$$G^k = \bigcup_{i \in I_h(F(x^k))} \{\nabla M(x^k) \nabla h_i(F(x^k))\}$$

where $I_h(F(x^k))$ is the set of indices for the piecewise affine parts h_i that define h and that are active at $F(x^k)$.

Generator set

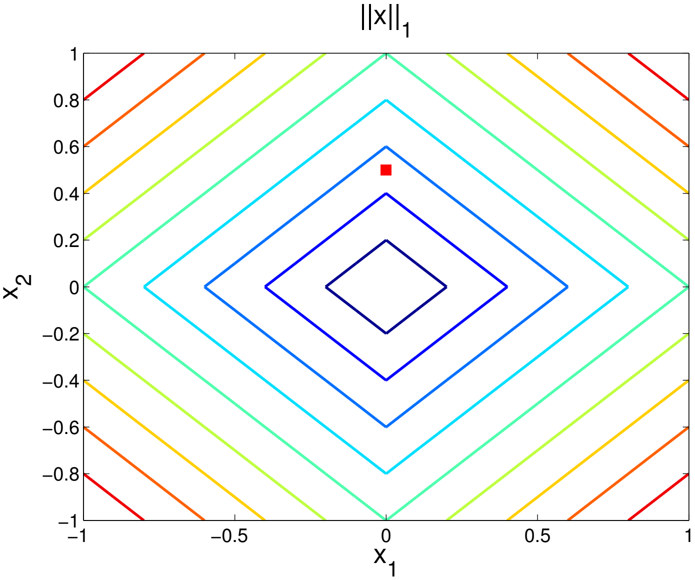
$$G^k = \bigcup_{i \in I_h(F(x^k))} \left\{ \nabla M(x^k) \nabla h_i(F(x^k)) \right\}$$

where $I_h(F(x^k))$ is the set of indices for the piecewise affine parts h_i that define h and that are active at $F(x^k)$.

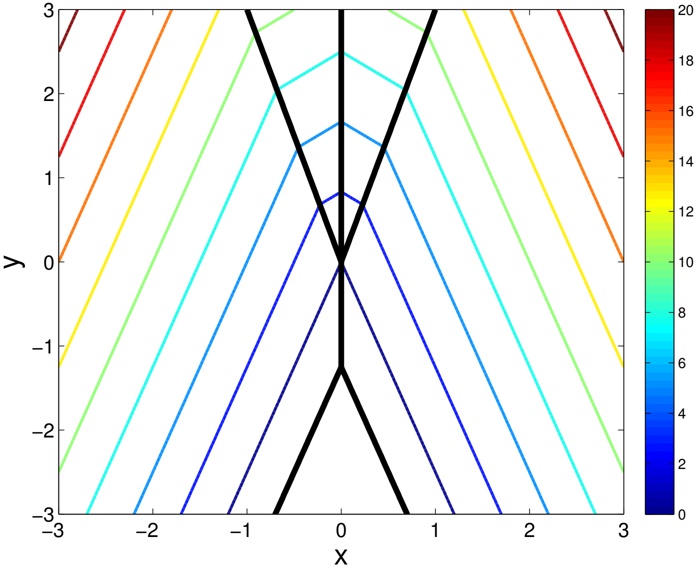
Or, given a set of points $Y = \left\{x^k, y^2, \ldots, y^p\right\} \subset \mathcal{B}(x^k, \Delta_k)$,

$$G_k = \bigcup_{y \in Y} \bigcup_{i \in I_h(F(y))} \left\{ \nabla M(x^k) \nabla h_i(F(y)) \right\}$$

Generator set



Generator set



Trust region methods

Composite nonsmooth case

1. Build a model $m_k^{F_i}$ of F_i at x^k.

2. Use ∇m^{F_i} to form $\nabla M(x)$ and build a set of generators G^k. Set ξ^k to be the minimum norm element in G^k.

Trust region methods

Composite nonsmooth case

1. Build a model $m_k^{F_i}$ of F_i at x^k.

2. Use ∇m^{F_i} to form $\nabla M(x)$ and build a set of generators G^k. Set ξ^k to be the minimum norm element in G^k.

3. Build a model of f at x^k with a gradient ξ^k and minimize that over $\mathcal{B}(x^k, \Delta_k)$ to obtain $x^k + s^k$ and evaluate $h(F(x^k + s^k))$.

Trust region methods

Composite nonsmooth case

1. Build a model $m_k^{F_i}$ of F_i at x^k.

2. Use ∇m^{F_i} to form $\nabla M(x)$ and build a set of generators G^k. Set ξ^k to be the minimum norm element in G^k.

3. Build a model of f at x^k with a gradient ξ^k and minimize that over $\mathcal{B}(x^k, \Delta_k)$ to obtain $x^k + s^k$ and evaluate $h(F(x^k + s^k))$.

4. Ensure the correct manifolds are in G^k, depending on the case. Either add them to G^k and go to 2, or calculate ρ_k.

Trust region methods

Composite nonsmooth case

1. Build a model $m_k^{F_i}$ of F_i at x^k.
2. Use ∇m^{F_i} to form $\nabla M(x)$ and build a set of generators G^k. Set ξ^k to be the minimum norm element in G^k.
3. Build a model of f at x^k with a gradient ξ^k and minimize that over $\mathcal{B}(x^k, \Delta_k)$ to obtain $x^k + s^k$ and evaluate $h(F(x^k + s^k))$.
4. Ensure the correct manifolds are in G^k, depending on the case. Either add them to G^k and go to 2, or calculate ρ_k.
5. If $\rho_k > \eta > 0$, $x^{k+1} = x^k + s^k$, $\Delta_{k+1} = \gamma_{\mathrm{inc}}\Delta_k$.
   Else $x^{k+1} = x^k$, $\Delta_{k+1} = \gamma_{\mathrm{dec}}\Delta_k$.

Conclusions

- Nonsmooth problems appear in many places and motivate active optimization research.
  See: **MS76** Wed. 10:30 AM – 12:30 PM, BCEC Room 253C:
  Sensitivity Analysis and Optimality Conditions in Nonsmooth Problems.

Conclusions

- Nonsmooth problems appear in many places and motivate active optimization research.
  See: **MS76** Wed. 10:30 AM – 12:30 PM, BCEC Room 253C: Sensitivity Analysis and Optimality Conditions in Nonsmooth Problems.

- Functions of the form $h(F(x))$ with piecewise linear h encompass a large class of nonsmooth f.

Conclusions

- Nonsmooth problems appear in many places and motivate active optimization research.
  See: **MS76** Wed. 10:30 AM – 12:30 PM, BCEC Room 253C: Sensitivity Analysis and Optimality Conditions in Nonsmooth Problems.

- Functions of the form $h(F(x))$ with piecewise linear h encompass a large class of nonsmooth f.

- Often, the nonsmoothness has a known form. Exploiting this can be beneficial to performance.
  Larson, Menickelly, Wild. "Manifold Sampling for L1 Nonconvex Optimization."
  See: **IT5** Wed. 8:30 AM – 9:15 AM, Grand Ballroom: Beyond the Black Box in Derivative-Free and Simulation Based Optimization.